DYNAMIC ECONOMETRIC MODELS
Vol. 6 – Nicolaus Copernicus University – Toruń – 2004

Tadeusz Kufel
Nicolaus Copernicus University in Toruń

General-to-Specific Modelling vs. Congruent Modelling in PcGets

1. Introduction

The aim of this paper is to compare the concept of general to specific modelling with the concept of congruent modelling and to indicate their similarities and differences. Moreover the purpose is to present the PcGets module of OxMetrics package which enables to select the best econometric model starting with the general (full) specification of a dynamic econometric model.

2. Different approaches to the specification of dynamic econometric models

Different approaches to the starting specification of a dynamic econometric model can be observed. Some approaches refer to causal relationships, others to internal structure but with omitting the causality issue and other combine the properties of the two. The concepts of general to specific modelling by Hendry[1] and congruent modelling by Zieliński[2] belong to the latter approach with regard to model specification.

Such an approach, i.e. using the economic theory to establish the causal structre and including some elements of process structure into the model, is not so new, because already at the beginning of the twenty century the examples of such studies could be found in the works of Jevons, Moore, Hooker, Pearson, Lange, to mention only the elimination of trend or seasonality components by

[1] See: e.g. Davidson, Hendry, Srba, Yeo (1978), Hendry (1995), Hendry (2000).
[2] See: e.g. Zieliński (1984), (1991), Talaga, Zieliński (1986).

subtracting them or including them into a model[3]. However such solutions resulted from the intuition of a researcher, but not from systematic analysis of processes structure.

The theoretical basis for such an approach was given by the Frisch-Waugh theorem, which indicates that the introduction of time variable *t* to the model eliminates the linear trend from all processes included in the model[4]. This theorem was generalized by Tintner (1952, p. 301) on the case of polynomial trend, and by Lovell (1963, pp. 993–1009) – on the case of seasonality. The next generalization was made by Stone (1962, pp. 401-403) on the case of any set of variables. Stone proved that the estimation results of a model including the given set of variables are the same as those obtained for processes from which the influence of that set of variables was eliminated. All these theorems show that the inclusion of the elements of internal process structure (trend, seasonality) eliminates those components from all variables included into a model. As a result the interpretation of model parameters should be referred to the relationship among stationary parts of given processes.

Other examples of improving the model specification were attempts of including the dummies to describe the outlying changes in the mean of process. Many different suggestions for using dummies in dynamic econometric models in order to take into account the changes in mean and/or in variance, changes in structural parameters, etc., can be found in Welfe (1977), pp. 95–115. As a result the variability of intercept could be better described. Therefore the inclusion of dummies should be treated as the starting point for making the harmonic structure of an endogenous process and jointed harmonic structure of explanatory processes and residual process congruent[5].

In the seventies of the twentieth century new trends in econometrics are being observed. These trends are related with the critique of existing modelling, i.e. modelling for economic random variables, but not for economic stochastic processes. The examples of such critiques can be found in Hendry (1974), (1977), 1980), (1984)[6]. He indicated that the newest methods and robust testing should be used by econometricians in time series analysis. This suggests that they found, in such a way, statistical relationships should be put before the economic theory. A similar view was expressed by Sims (1980), who preferred to build models without imposing too many assumptions and to treat all processes as endogenous (VAR modelling).

The critics of traditional approach to econometric modelling pointed to excessive exploitation of samples for obtaining the required results. In Marchi and

[3] For more information about the use of the structure of processes in different concepts of econometric modelling see: Kufel (2002), ch. 1.

[4] For more details see: Hozer, Zawadzki (1990), pp. 11–31.

[5] See: Zieliński (1984), p. 46.

[6] The collection of many articles referring to those critics can be found in Hendry (1993).

Gilber (1989), p. 123, was underlined that the traditional econometrics seems to be the caricature of simple to general modelling.

However, the opinion that in the seventies of the twentieth century, the crisis in econometrics happened is not justified, because this period should be treated rather as a starting point for the development of new trends in dynamic econometric modelling on the basis of the theory of stochastic processes, where the dynamic and stochastic specifications are jointly taken into account (see: Hendry, Pagan, Sargan (1984), p. 1025).

The first information about the congruence idea can be found in Granger (1981), where the congruence idea was explained first by presenting the non-congruent model of the following form[7]

$$y_t = a + bx_t + cz_t + e_t, \tag{1}$$

where y_t stands for the seasonal process, x_t and z_t – for the nonseasonal processes, e_t – for the white noise process.

Granger pointed that the model (1) would never have the required properties of residual process e_t, i.e. white noise properties, because the non-seasonal processes x_t and z_t cannot describe the seasonal component existing in y_t. As a result unexplained seasonal changes in y_t will occur in e_t, so the residual process e_t will contain the seasonal component.

The observation that the data are not "consistent" with the assumptions accepted in the model was the starting point for developing the concept of dynamic congruent modelling which takes into account the information about the internal structure of studied processes.

The concept of dynamic congruent modelling is Profesor Zygmunt Zieliński (1984)[8]. In his paper on the time variability of structural parameters he outlined this concept. The generalization of this concept can be found in Zieliński (1985a), (1985b), (1991), Talaga, Zieliński (1986), ch. V.

The model congruence in Zieliński sense is understood as the congruence of the harmonic structure of an endogenous process and jointed harmonic structure of explanatory processes and residual process, which is independent of explanatory processes. The model, for which all taken processes are white noises, is always congruent. This model has the form

$$\varepsilon_{yt} = \sum_{i=1}^{k} \rho_i \varepsilon_{x_i t} + \varepsilon_t. \tag{2}$$

[7] See: Granger (1981), p. 122.

[8] The paper was submitted to the editorial office in November 1982.

Model (2) is congruent, because the harmonic structures of the left and the right side of the equation are identical or similar, in other words, their spectra are parallel to frequency axis.

Let Y_t and X_{it} ($i = 1, .., k$) denote the endogenous process and an explanatory processes vector respectively, for which the fundamental model describing their internal structure are as follows:

– models describing non-stationary components

$$Y_t = P_{yt} + S_{yt} + \eta_{yt}, \qquad X_{it} = P_{x_i t} + S_{x_i t} + \eta_{x_i t}, \tag{3}$$

where $P_{yt}, P_{x_i t}$ stand for polynomial function of time variable t for respective processes, $S_{yt}, S_{x_i t}$ – seasonal components with constant or changing amplitude of fluctuations for respective processes, $\eta_{yt}, \eta_{x_i t}$ – stationary autoregressive processes respectively.

– autoregressive models

$$B(u)\eta_{yt} = \varepsilon_{yt}, \qquad A_i(u)\eta_{x_i t} = \varepsilon_{x_i t}, \tag{4}$$

where $B(u)$, $A_i(u)$ – stationary autoregressive backshift operators for which all roots of equations $|B(u)| = 0$ i $|A_i(u)| = 0$ lie outside the unit circle, $\varepsilon_{yt}, \varepsilon_{x_i t}$ – white noises for respective processes.

Information about the internal structure of all studied processes enables the construction of the dynamic congruent model on the basis of the relationship for white noises described by model (2).

The congruent model for real processes, i.e. Y_t and X_{it}, is obtained by the following substitution: the white noises in model (2) are replaced with those from model (4), next the autoregressive processes $\eta_{yt}, \eta_{x_i t}$ are replaced with those obtained from model (3). After further transformations the model takes the form:

$$B(u)Y_t = \sum_{i=1}^{k} A_i^*(u)X_{it} + P_t + S_t + \varepsilon_t. \tag{5}$$

In model (5) the residual process ε_t is the same as in model (2). This means that the congruence condition of harmonic structures for both sides of the equation was satisfied. The congruent model (5) contains all elements of internal structure of given processes.

The above outlined concept of dynamic congruent modelling points out that the information about the internal structure of studied processes should be necessarily taken into account on the specification model stage. Building model on the basis of the equation (2) for white noises ensures that the congruence condition in such a specified model will be satisfied and any residual process will have white noise properties. Worth noting is the fact that this is known already on the stage of model construction.

The general-to-specific modelling initiated by Hendry[9] means the construction of dynamic model starting with a large, general model without restrictions and finishing with the model reduced by the use of statistical tests. As the starting (general) model is taken the autoregressive distributed lag model (ADL). The one-equational ADL model has the following form:

$$B(u)Y_t = \sum_{i=1}^{k} A_i(u)X_{it} + \varepsilon_t, \tag{6}$$

where $B(u)$ and $A_i(u)$ are autoregressive polynomials each of m-order.

For simplification the models with the same autoregression order m for all studied processes are built. Then the model is denoted as ADL(m).

The modified ADL model, ADL(m_0, m_1, ..., m_k), in which each studied process can have different order of autoregression has the form

$$\sum_{j=0}^{m_0} \beta_j Y_{t-j} = \sum_{j=1}^{k} \sum_{i=0}^{m_k} \beta_{ji} X_{jt-i} + \varepsilon_t. \tag{7}$$

The ADL model of the form (7) assuming the stationarity condition and correct specification of lags are satisfied can be treated as the dynamic congruent model. Hence for stationary processes the model (7) is the specific case of a full (starting version) congruent model. However the stationarity assumption for real economic processes is difficult to be kept, and then the model (7) is not sufficient.

Lately, particularly after the paper by Hoover and Perez (1999), the procedure of building the one-equational dynamic model has been developed[10]. This procedure consists of building the starting extended specification of a model, the so-called general unrestricted model, GUM, and reducing it to final model, so called congruent empirical model (see: Hendry (2000), p. 467). The difference in building the model according to general to specific modelling by Hendry and dynamic congruent modelling by Zieliński resolves itself to that in the former that there is the lack of strict specification of the internal structure of studied processes for the starting model. Therefore, the representatives of Lon-

[9] See: Davidson, Hendry, Srba, Yeo (1978).

[10] See: Hendry, Krolzig (1999), Krolzig, Hendry (2000).

don School of Economics, LSE[11], recommended adding into the model lagged processes as long as the residual process will have white noise properties. This means that the general-to-specific modelling indirectly uses the information about the structure of studied processes to build the starting model specification.

3. PcGets – module of OxMetrics[12]

PcGets is the software, which enables the automatic selection of a final dynamic econometric model starting with the GUM or full model. To build the empirical congruent model PcGets uses the **ge**neral-**t**o-**s**pecific modelling.

The procedure of building the econometric model consists of three stages:

A. Specification of GUM model – model formulation

The specification of a starting model is carried out by selecting: the endogenous process and its lags, explanatory processes and their lags, deterministic components (trend, seasonal dummies) – see: Fig. 1. There is only one condition to enable the model estimation, i.e. the number of processes cannot be larger than the number of observations. To satisfy the congruence condition the GUM model has to be specified in such a way that the residual process has white noise properties.

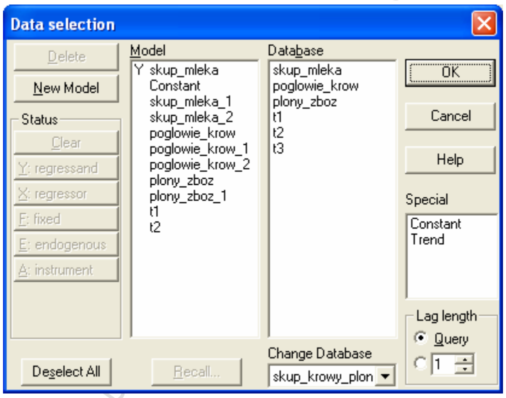

Fig. 1. Specification of full model (GUM)
Source: Window of PcGets.

The model constructed according to the concept of congruent modelling satisfies all conditions of the GUM, and additionally has a precisely determined procedure of selecting the lags' number and including deterministic

[11] The outline of the LSE methodology is described in Granger (ed.) (1990), pp. 279–364, Hendry (1995).

[12] PcGets was developed by Hendry, Krolzig (2001), and OxMetrics – by Doornik (2001). For more details see the attached materials at the end of this book.

components. The concept of congruent modelling can be applied in the case of integrated processes, but then the number of lags should be extended[13].

B. Selection of tests and strategies of the elimination of insignificant processes (selection of significance level) and types of reports

The application of different tests is recommended to select the empirical (final) congruent model. These tests and reasons for using them were broadly depicted in Hendry, Krolzig (2001, pp. 101–169). When using the PcGets to automatically select final model the choice of significance level is important. The package suggests three strategies, i.e. liberal, conservative and additionally user strategy with significance levels established by the user.

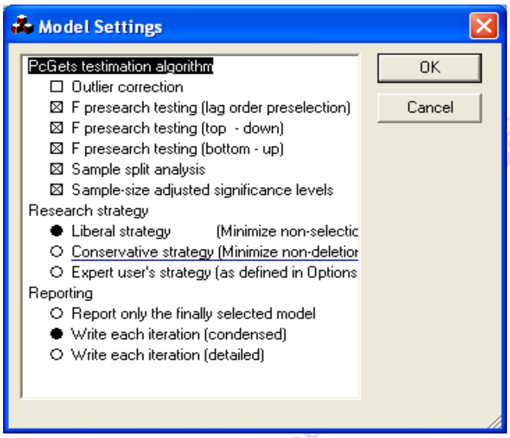

Fig. 2. Selection of tests, strategies (significance levels) and reports' types
Source: Window of PcGets.

However it should be noted that the significance level cannot be chosen arbitrarily. It results from that in large samples (n µ 60) for recovering the data generating model the significance level lower than 5% is required, and in small samples – markedly higher than 5%. The appropriate significance level with regard to sample size can be chosen by calculating the following formula: $\alpha = 1.6\, n^{-0.9}$ or $\alpha = n^{-0.8}$. The different strategies for selecting the significance levels are displayed at Fig. 3.

The liberal strategy suggests for small samples ($n < 60$) selecting the 10% significance level, which is lowered in the large sample ($n > 1000$) to the 1%. While the conservative strategy suggests choosing the 5% significance level in small samples, and 0.1% – in large samples (see: Hendry, Klozig (2001), p. 197).

[13] See: Zieliński (1995), Piłatowska (2003).

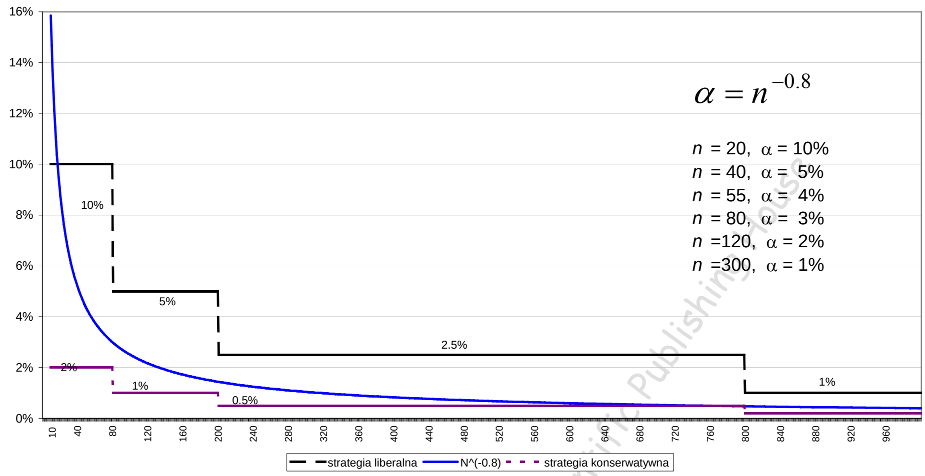

Fig. 3. Strategies for selecting the significance levels with regard to sample size

C. Selection of estimation methods

Figure 4 presents the window for selecting the estimation methods and sample range.

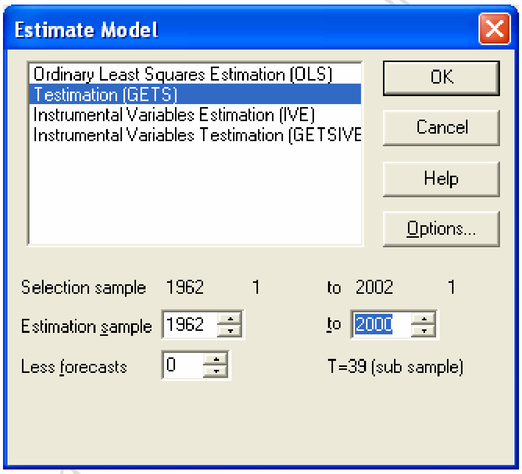

Fig. 4. Selection of estimation methods and sample range
Source: Window of PcGets.

PcGets makes it possible to estimate a model only by the ordinary least squares method and instrumental variables method (see: Fig. 4). Selecting the option *Testimation* (GETS/GETSIVE) runs the algorithm for automatic elimination of insignificant processes (excessive processes) which ends with the choice of the final (empirical) congruent model having the required properties. PcGets displays results in the form of a text report, generates automatically the set of figures to evaluate the fit of the model, properties of residuals and their squares (acf, pacf, spectrum) and forecasts with errors (*ex ante* and *ex post*).

4. Final remarks

Since 1984 many dynamic congruent models have been built during the different research projects realized in the Department of Econometrics and Statistics. The final version of a congruent model with white noise properties of residuals is obtained after the reduction of insignificant processes. Most frequently the *a posteriori* selection method is used to eliminate insignificant processes. But in small samples its use may give the excessive elimination leading to the loss of white noise properties of residuals. In that case another selection method, e.g. all possible regressions, should be used.

It is seemed that the automatic selection procedure carried out in PcGets, which assumes the inclusion of earlier eliminated processes, can be thought as an effective tool in selecting such a specification that the white noise properties of a residual process are held.

References

Davidson, J. H., Hendry, D. H., Srba, F. I, Yeo, S. (1978), Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in United Kingdom, *The Economic Journal*, vol. 88, 661–692.

Doornik J. A. (2001), *Ox. An Object-Oriented Matrix Language*, London, Timberlake Consultants Press, fourth edition.

Frisch, R., Waugh F. (1933), Partial Time Regressions as Compared with Individual Trends, *Econometrica*, vol. 1, 387–401.

Granger, C. W. J. (1981), Some Properties of Time Series Data and Their Use in Econometric Model Specification, *Journal of Econometrics*, vol. 16.

Granger, C. W. J. (ed.) (1990), *Modelling Economic Series. Readings in Econometric Methodology*, Oxford University Press.

Hendry, D. F. (1980), Econometrics: Alchemy or Science?, *Econometrica*, vol. 47, 387–406.

Hendry, D. F. (1993), *Econometrics: Alchemy or Science? Essays in Econometric Methodology*, Blackwell, Oxford UK, Cambridge USA.

Hendry, D. F. (1995), *Dynamic Econometrics*, Oxford University Press.

Hendry, D. F. (2000), *Econometrics: Alchemy or Science? New Edition*, Oxford: Oxford University Press.

Hendry, D. F., Krolzig, H. M. (1999), Improving on 'Data Mining Reconsidered' by K. D. Hoover and S. J. Perez, *The Econometrics Journal*, vol. 2, no. 2, 202–219.

Hendry D. F., Krolzig H. M. (2001), *Automatic Econometric Model Selection Using PcGets 1.0*, Published & Distributed by Timberlake Consultants Ltd., London.

Hendry, D. F., Pagan, A. R., Sargan, J. D. (1984), *Dynamic Specification,* w: *Handbook of Econometrics*, vol. II, Elservier Science Publishers, 1023–1100.

Hoover, K. D., Perez S. J. (1999), Data Mining Reconsidered: Encopassing and the General–to–Specific Approch to Specification Search, *The Econometrics Journal*, vol. 2, no. 2, 167–191.

Hoover, K. D., Perez S. J. (1999), Reply to our Discussants, *The Econometrics Journal*, vol. 2, no. 2, 244–247.

Hozer, J., Zawadzki, J. (1990), *Zmienna czasowa i jej rola w badaniach ekonometrycznych (Time Variable and Its Role in Econometric Analysis)*, PWN, Warszawa.

Krolzig, H. M., Hendry, D. F. (2000), Computer Automation of General–to–Specific Model Selection Procedures, *Journal of Economic Dynamics and Control*, vol. 25, 831–866.

Kufel, T. (2002), *Postulat zgodności w dynamicznych modelach ekonometrycznych (Congruence Postulate in Dynamic Econometric Models)*, Wydawnictwo UMK Toruń.

Lovell, M. C. (1963), Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis, *Journal of American Statistical Association*, no. 58, New York.

Piłatowska, M. (2003), *Modelowanie niestacjonarnych procesów ekonomicznych. Studium metodologiczne (Modelling Non-Stationary Economic Processes. Methodological Study)*, Wydawnictwo UMK, Toruń.

Talaga, L., Zieliński, Z. (1986), *Analiza spektralna w modelowaniu ekonometrycznym (Spectral Analysis in Econometric Modelling)*, PWN, Warszawa.

Tintner, G. (1952), *Econometrica,* John Wiley & Sons, New York.

Welfe, W. (red.), (1977), *Metody ekonometryczne (Econometric Methods)*, t. I, PWE, Warszawa.

Wiśniewski, J. W., Zieliński, Z. (1985), *Ekonometria,* cz. I, Wydawnictwo UMK, Toruń.

Zellner, A., Palm, F. (1974), Time Series Analysis and Simultaneous Equation Econometric Models, *Journal of Econometrics*, vol. 2.

Zieliński, Z. (1984), Zmienność w czasie strukturalnych parametrów modelu ekonometrycznego, (Time Variablility of Structural Parameters in Econometric Model), *Przegląd Statystyczny (Statistical Survey)*, z. 1/2.

Zieliński, Z. (1985a), Liniowe modele opisujące zależności stacjonarnych procesów ekonomicznych (Linear Models Describing Relationships of Stationary Economic Processes), in: Wiśniewski, J. W., Zieliński, Z. (1985), *Ekonometria (Econometrics)*, part I, Wydawnictwo UMK, Toruń, 277–315.

Zieliński, Z. (1985b), Liniowe modele opisujące zależności niestacjonarnych procesów ekonomicznych (Linear Models Describing Relationships of Non-stationary Economic Processes), in: Wiśniewski, J. W., Zieliński, Z. (1985), *Ekonometria (Econometrics)*, part I, Wydawnictwo UMK Toruń, 316–346.

Zieliński, Z. (1991), *Liniowe modele ekonometryczne jako narzędzie opisu i analizy przyczynowych zależności zjawisk ekonomicznych (Linear Econometric Models as a Tool of Description and Analysis of Causal Realationships among Economic Phenomena)*, Wydawnictwo UMK, Toruń.